Co-Relations between Speeding and Crashes - City of Chicago

Andrew Weathers, Madhumita Krishnan
Siddhant Aggarwal

Goal

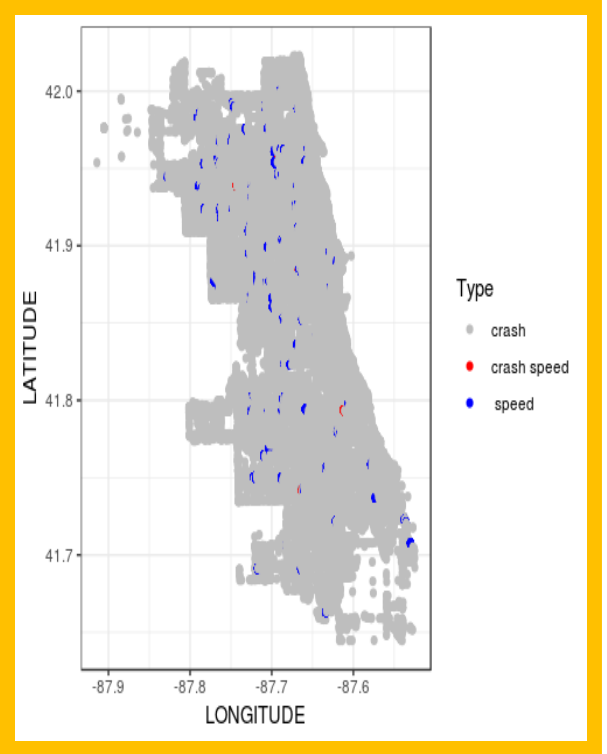
Our goal is to determine relationship between speeding violations, as detected by cameras at intersections, and number of crashes in the city of Chicago.

Dataset

We use two datasets. The first one lists over 220,000 traffic crashes and the second lists over 150,000 speeding violations captured across the city of Chicago. The speeding violations have been captured by 150 cameras. Both the datasets span over a period of 2014-2018.

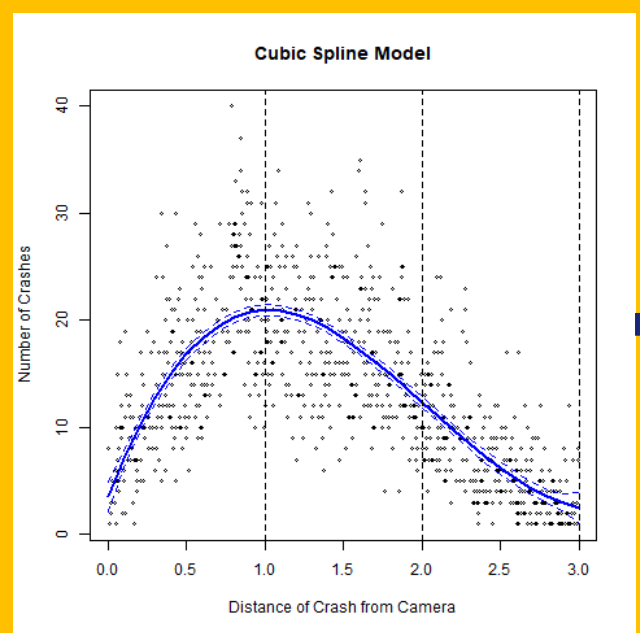
Models Used

Linear Regression, Non-Linear models like Polynomial regression, Smoothing splines, Regression Splines



Crashes vs Distance from Speed Cameras

Using Linear and other Non-Linear regression techniques, we attempted to predict a relation between a crash and its distance from a speed detection camera.
Using Regression splines we could predict how far a crash occurred from a camera (critical distance at 1 Km).
We make a prediction for 1Km range around speeding cameras. It is predicted that, on an average, 126 crashes shall take place per day.



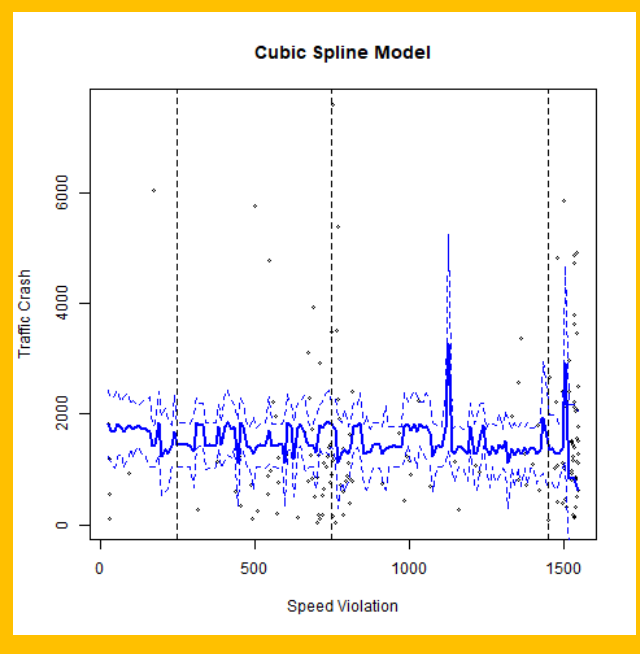

```
Call:
lm(formula = count ~ bs(value, knots = knot.position), data = Train)

Residuals:
    Min      1Q  Median      3Q     Max
-13.309  -3.287  -0.690   2.645  31.969

Coefficients: (1 not defined because of singularities)
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        3.7711     0.6595   5.718 1.23e-08 ***
bs(value, knots = knot.position)1 12.1169     1.2454   9.729  < 2e-16 ***
bs(value, knots = knot.position)2 21.2499     0.8096  26.248  < 2e-16 ***
bs(value, knots = knot.position)3  8.7762     1.1275   7.783 1.11e-14 ***
bs(value, knots = knot.position)4 -0.3613     0.9220  -0.392    0.695
bs(value, knots = knot.position)5 -1.0254     0.9821  -1.044    0.297
bs(value, knots = knot.position)6       NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.272 on 2063 degrees of freedom
Multiple R-squared:  0.5581,    Adjusted R-squared:  0.5571
F-statistic: 521.2 on 5 and 2063 DF,  p-value: < 2.2e-16
```

Speed Violation vs Traffic Crashes

The results for Linear Regression using Lasso were found to be

- MSEImprovement- 26.76%
- lassoMSE- 37090.46
- nullMSE- 50647.48

We used Non-Linear models and found using R-squared values and p-values that the co-relation was insignificant.
<u>Using Cubic Spline Model, we found the co-relation to be maximal but still insignificant.</u>



Future Work

- We want to establish a more generalized co-relation technique where we can run the same script across other datasets which include common features like Longitude, Latitude, Date etc.
- We also would like to exploit additional datasets for cities like Seattle and New York city to find similarities.